Multimodal Objective Sensing to Assess Individuals with Context (MOSAIC) Proposers' Day

**Alexis Jeannotte, Ph.D.**
August 2, 2016

MOSAIC Overview

- MOSAIC is anticipated to be a multi-year research and development program.

- The program seeks to develop and validate unobtrusive sensor-based methods to assess static and dynamic aspects of an individual's cognitive abilities and styles, interpersonal style, and well-being.

- The MOSAIC program aims to take advantage of multimodal mobile, worn, and carried sensors, as well as other media sources, to aggregate activity and behavior, physiology, social dynamics, physical location and proximity, as well as other novel features to develop personalized assessments.

Background
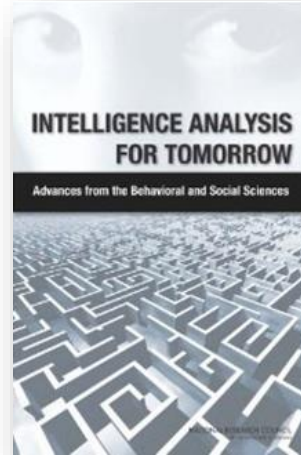
"Use Scientific Methods for Workforce Development"

"Ensure that IC agencies use evidence-based methods to recruit, select, train, motivate, and retain an adaptive workforce able to achieve the performance levels required by IC missions."

The intelligence community (IC) operates in an increasingly complex, turbulent, and fast-changing threat environment (see Fingar, 2011). This environment has important implications for recruiting, training, organizing, retaining, and managing the IC workforce. In stable environments, organizations can rely on stable work practices overseen by a relatively rigid administration that directs a generally hierarchical and compliant workforce. In turbulent environments, organizations require innovative work practices, flexible administration, and a creative, inventive workforce, given rein to find new approaches and novel solutions.

-Intelligence Analysis for Tomorrow, National Academies Press, 2008
http://www.nap.edu/openbook.php?record_id=13040&page=3

Background - Overview

- Evaluating, supporting and maintaining our workforce throughout the length of their career is an important and persistent need across the IC.

- There are special considerations for the national security workforce, to include:

  - High tempo and high stress environments
  - Complex and multidimensional problem sets
  - Maintaining peak cognitive performance & decision-making skills
  - Maintaining motivation of self and colleagues
  - Psychological, cognitive, and behavioral consistency or flexibility
  - High trustworthiness and integrity

- Evaluations of these aspects occur before and during the hiring process, as well as throughout someone's employment tenure.

Background – Today's Challenges
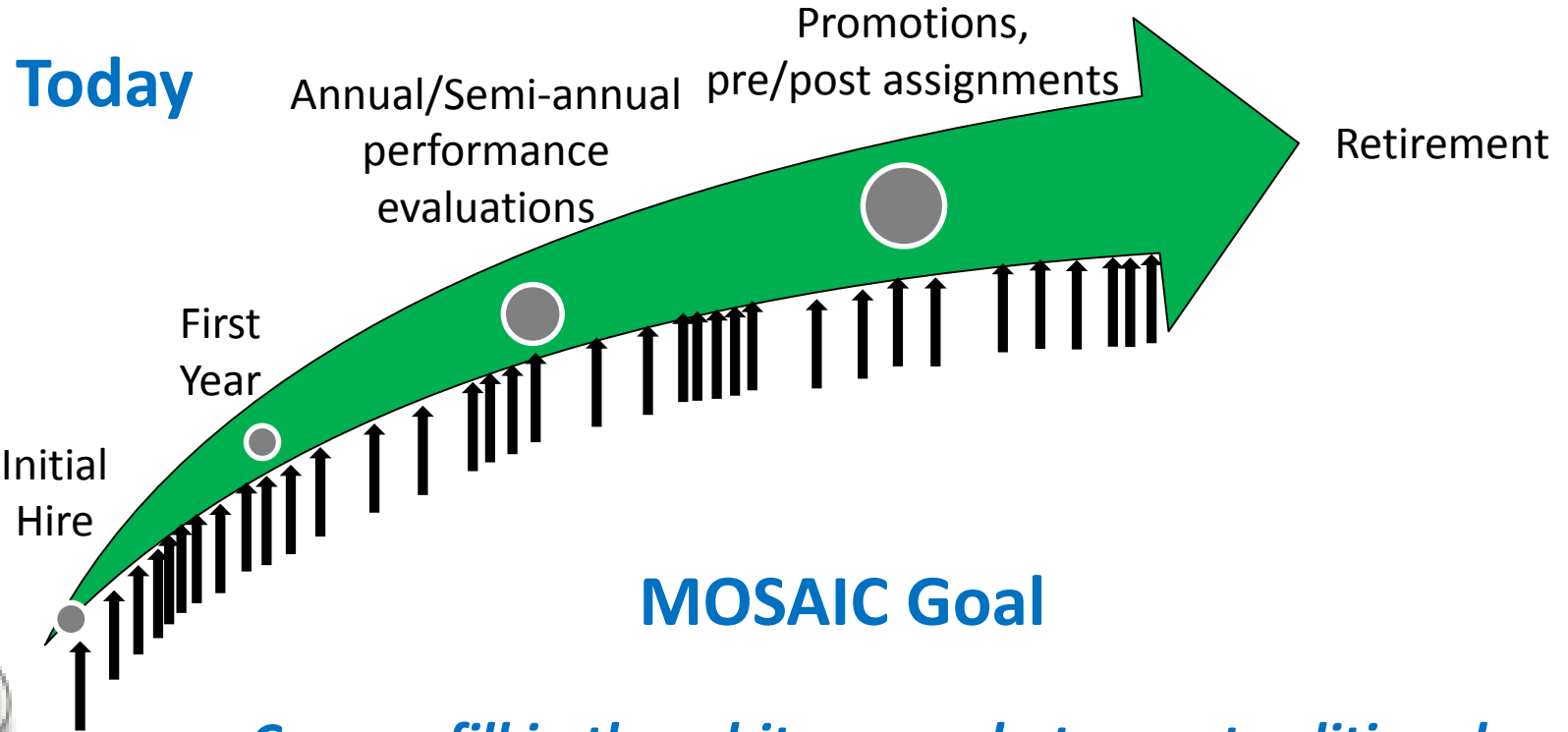
Existing Evaluation Tools

- Standardized testing
- Self- report & questionnaires
- Cognitive tasks
- Interview
- Assessment centers
- Situational judgment tests

Limitations

- Controlled or unnatural testing environment
- Limited accounting of the context
- One-time assessment or long intervals between evaluations
- Concerns about validity of retesting
- Potential for biases stemming from impression management or socially desirable responding
- Potential for evaluator bias

Background - Today

Today

Promotions, pre/post assignments

Annual/Semi-annual performance evaluations

Retirement

First Year

Initial Hire

MOSAIC Goal

Can we fill in the white space between traditional evaluation points with non-diagnostic evaluations of predictors and dimensions of job performance?

Background - Job Performance

"The total expected value to the organization of the discrete behavioral episodes that an individual carries out over a standard period of time"

– Motowidlo & Kell, 2013

Core dimensions of job performance (Viswesvaran et al., 2000; Rotundo & Sackett, 2002)

- **Task performance:** Proficiency in activities recognized as part of the job

- **Organizational citizenship behavior:** Discretionary behavior that in aggregate promotes effective functioning of the organization

- **Counterproductive behavior:** Voluntary behavior that significantly violates organizational norms, threatening well-being of an organization and/or its member.

- **Adaptability:** A supplemental dimension for high-stress/high-tempo jobs, e.g.:

Handling crisis situations	Interpersonal adaptability
Handling work stress	Cultural adaptability
Dealing with uncertain work situations	Learning new technologies & procedures

Background – Evaluating Job Performance

MOSAIC aims to look at both predictors and dimensions of job performance.

Predictors

Cognitive Abilities & Styles

Interpersonal Style

Well-Being

Dimensions

Task Performance

Organizational Citizenship Behavior

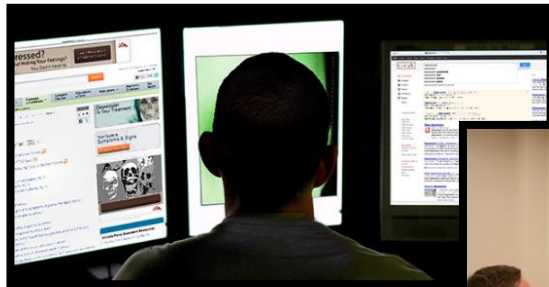Counterproductive Behavior

Adaptability

Background
Multimodal Sensing to Evaluate the Workforce

DARPA — DEFENSE ADVANCED RESEARCH PROJECTS AGENCY

Detection and Computational Analysis of Psychological Signals (DCAPS)

Mr. Wade Shen

http://www.darpa.mil

http://medvr.ict.usc.edu/projects/dcaps/

Harvard Business Review
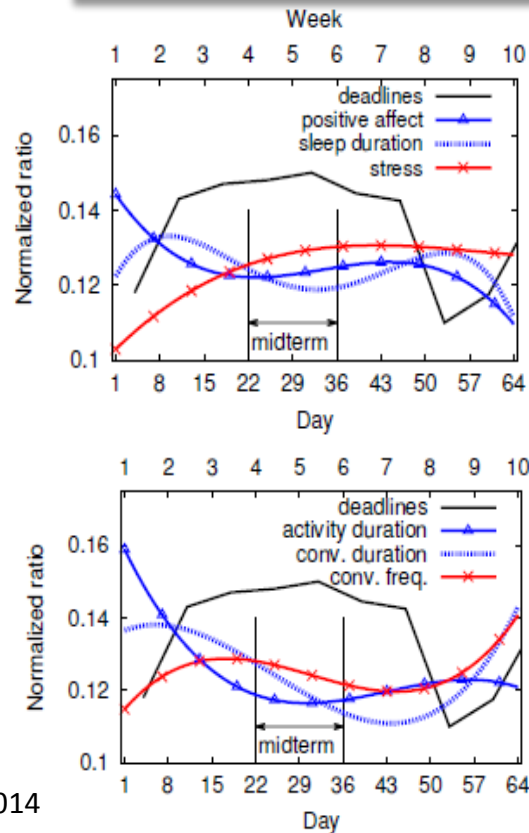
TECHNOLOGY

Wearables in the Workplace

by H. James Wilson

FROM THE SEPTEMBER 2013 ISSUE

"About 90% of companies now offer wellness programs, some of which encourage employees to use Fitbit and other devices that measure the quantity and intensity of their workouts and to employ simple visual and motivational tools to track their progress and help sustain their engagement. Because the programs are administered by third-party providers, employers can't see any individual's metrics. But the aggregate analytics give them robust insights about correlations between wellness, job satisfaction, and financial performance. The wellness program provider Carewise, whose members use Fitbit, has found that the health care costs of highly engaged participants rise just 0.7% annually, compared with 24% for less engaged participants."

Background
Detecting Stress from Daily Measurement

Stress Recognition using Wearable Sensors and Mobile Phones

Akane Sano
Massachusetts Institute of Technology
Media Lab
Affective Computing Group
75 Amherst Street, Cambridge,
MA, USA
akanes@media.mit.edu

Rosalind W. Picard
Massachusetts Institute of Technology
Media Lab
Affective Computing Group
75 Amherst Street, Cambridge,
MA, USA
picard@media.mit.edu
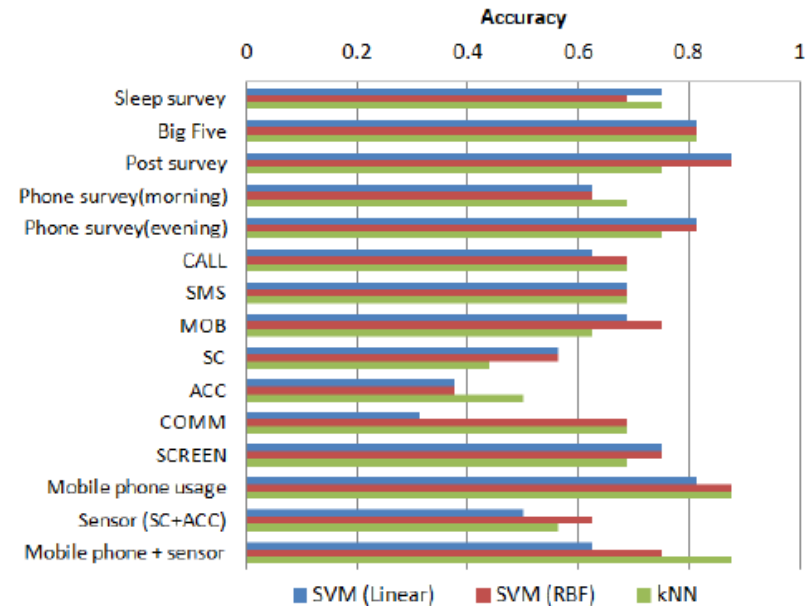
Fig. 4 Classification results using selected best features

Wang et al. 2014

Background
Context Based Assessments Aid Interpretation & Accuracy
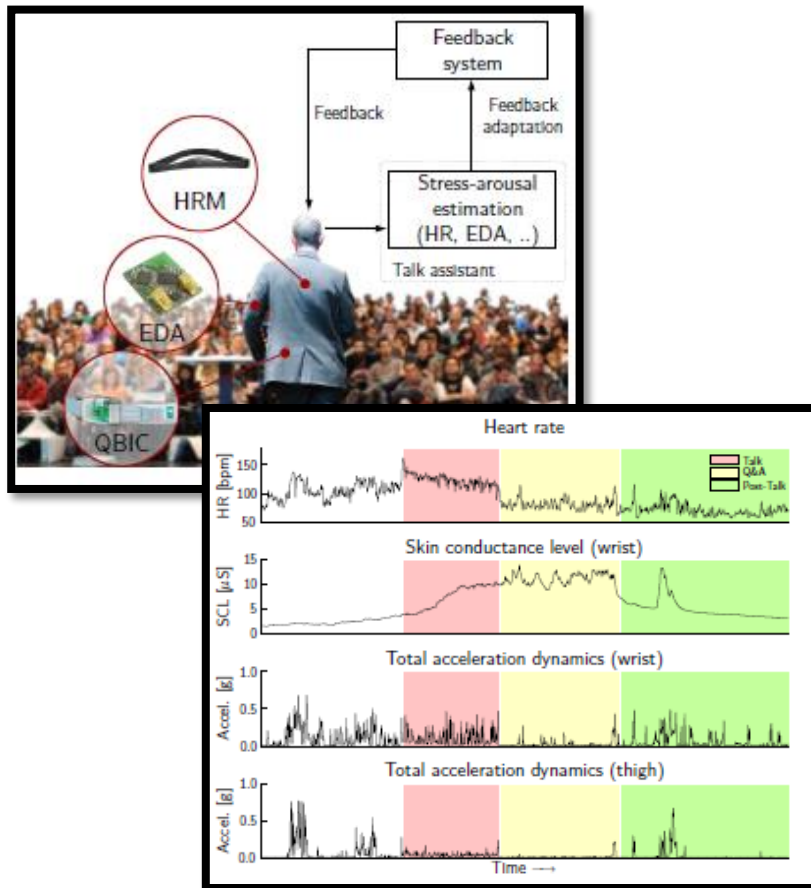
ETH Zurich, 2014

Towards Accurate Non-Intrusive Recollection of Stress Levels Using Mobile Sensing and Contextual Recall

Tauhidur Rahman[1], Mi Zhang[1], Stephen Voida[2], and Tanzeem Choudhury[1]

[1] Department of Information Science, Cornell University
[2] Department of Human-Centered Computing, Indiana University at IUPUI
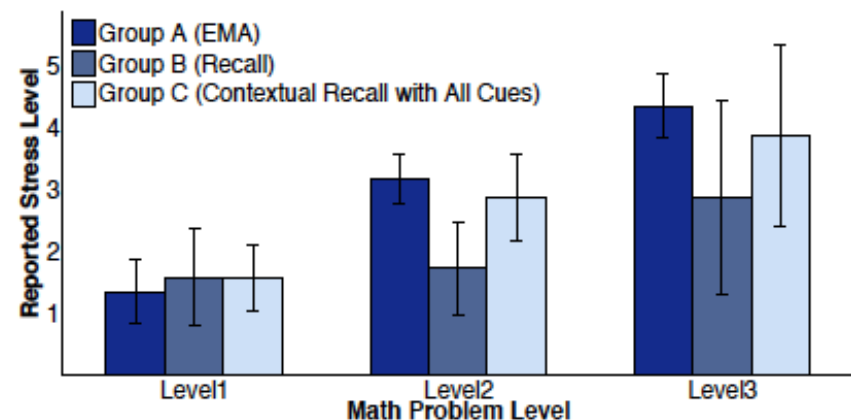
Background
Personalized Assessments Enable
More Accurate Understanding
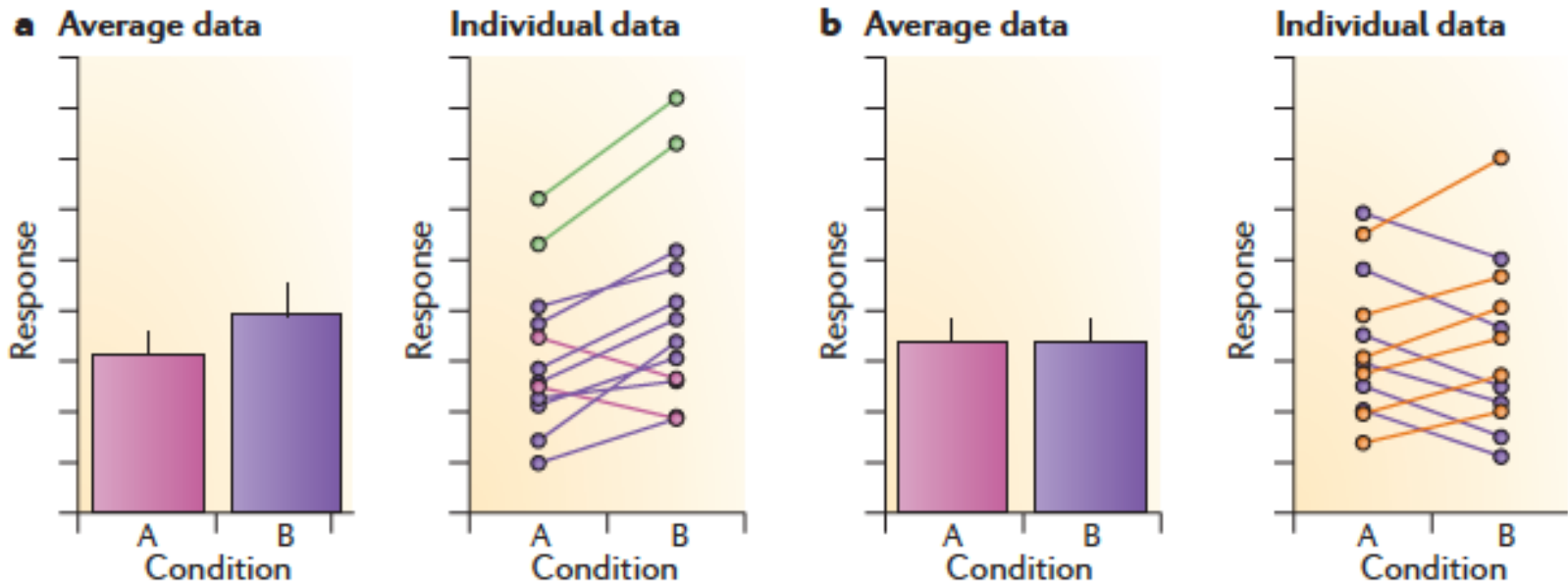
Kanai & Rees, 2011

Limitations of Previous Approaches for Real World Application

- Small study populations

- Use of students, friends/family as research subjects

- Single outcome measures or one outcome measure per construct

- Focus on the subject without context

- Group based analysis

- Time consuming data alignment and correction

Advance sensing, data analytics, and modeling with careful validation against ground truth

Persistent, unobtrusive, & passive sensing

Collect & Analyze Sensor Data

Validate Predictors & Dimensions of Job Performance Using Ground Truth Assessments

Cleaning

Processing

Modeling

Analysis

Cognitive Abilities & Styles

Interpersonal Style

Well-Being

Dimensions of Job Performance

Key Technical Challenges

- Bring the research and development of assessments out of the lab

- Develop methods for continuous or semi-continuous evaluation

- Move to unobtrusive methods that don't require continuous input from the subject

- Develop capabilities to capture the context, to help interpret subject-focused sensing

- Build capabilities to reduce the time and manpower required to handle, process, and analyze multimodal sensor data

- Advance techniques for personalized modeling that can capture variance over time

General Performer Research
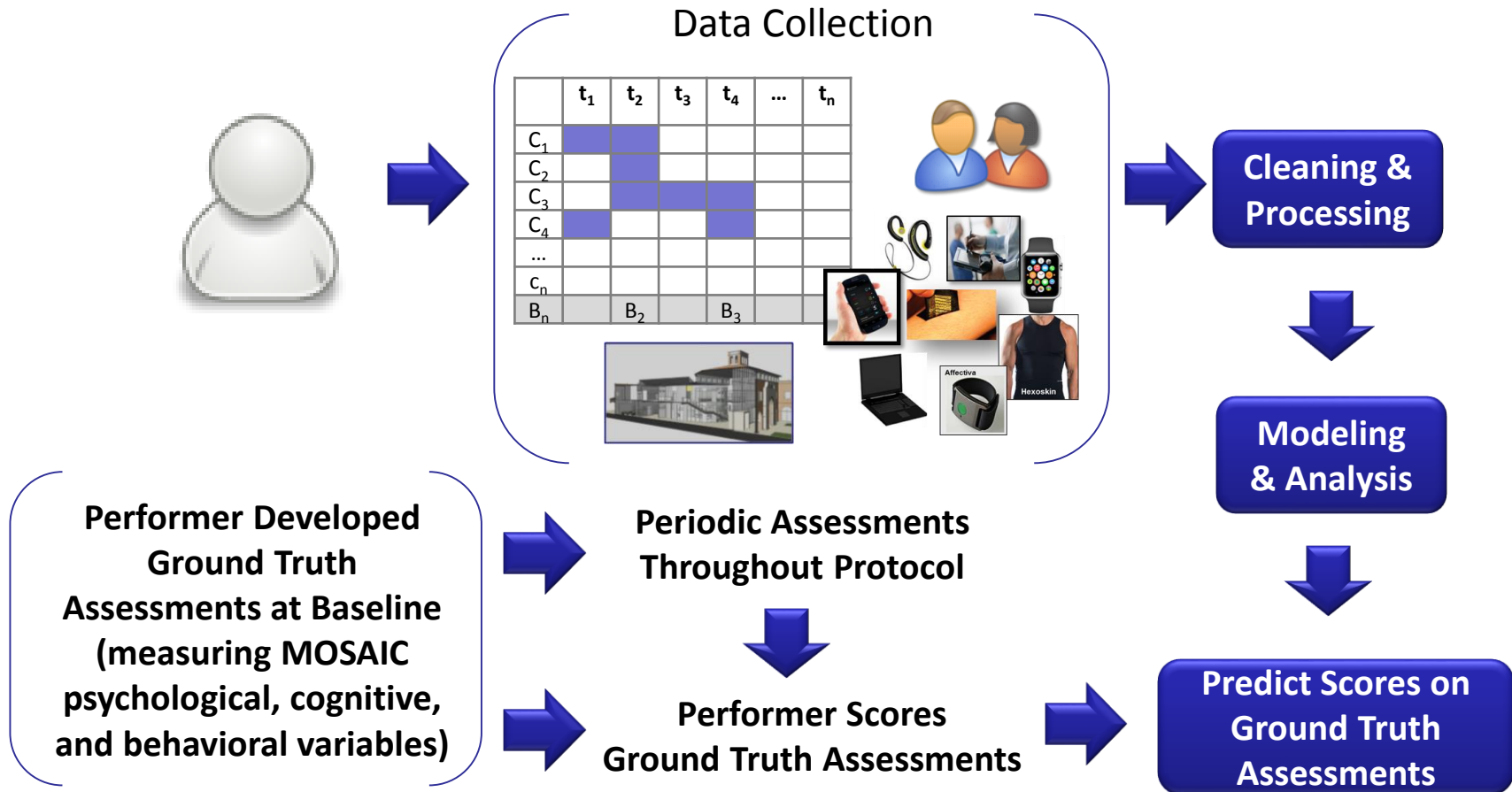
Data Collection

Cleaning & Processing

Modeling & Analysis

Predict Scores on Ground Truth Assessments

Performer Developed Ground Truth Assessments at Baseline (measuring MOSAIC psychological, cognitive, and behavioral variables)

Periodic Assessments Throughout Protocol

Performer Scores Ground Truth Assessments

General T&E Research

Replicate Performer Data Collection with New Subjects

	t_1	t_2	t_3	t_4	...	t_n
C_1						
C_2						
C_3						
C_4						
...						
c_n						
B_n		B_2		B_3		

Prep Raw Sensor Data

T&E provide unlabeled sensor data to performers

Performers submit:
- **Results on T&E Ground Truth Assessments**
- **All algorithms for independent verification**

T&E Developed Ground Truth Assessments at Baseline (measuring MOSAIC psychological, cognitive, and behavioral variables)

Periodic Assessments Throughout Protocols

T&E Scores Ground Truth Assessments

T&E scores & ranks performer results

Data Collection - Overview

Offerors will be asked to develop, propose, and refine a protocol to implement and test the proposed sensor(s), collect data from human subject volunteers,* and use this dataset to develop and advance processing, modeling, and analytic techniques.

Solutions may be:

- A mix of assessment during daily living and everyday activities combined with more structured evaluation periods

- Unobtrusive and passive data collection (e.g. mobile sensor) combined with active subject responses (e.g. phone survey, lab test)

- Offerors will be expected to develop and administer their own set of assessments of job performance predictors and dimensions

* Alternatively, may opt to use existing data sets or test strictly in the lab, but such a proposal must include a strong technical justification and the team still must provide T&E with a testing protocol that will enable T&E testing and data collection for the end of phase evaluation

Data Collection - Subjects

MOSAIC performers will be required to recruit, enroll, and test their own subject cohorts.

- Subject cohorts should be diverse: range of education, age, race, ethnicity, culture, socioeconomic status, experience, etc.

- Avoid focusing on undergraduate populations

- Power analyses will be necessary to justify proposed cohort sizes

- Estimates of cohort sizes should account for attrition, which should not be less than 25% (may be more)

- All enrolled subjects will be required to sign informed consent documents

Data Collection – Subject Privacy & Protections

MOSAIC will only fund research that emphasizes and employs best practices for subject safety, privacy, and protections.

- Offerors will be required to develop and propose methods that are safe, ethical, and can be approved by human subjects Institutional Review Boards (IRB)

- IRB protocols must stipulate that de-identified data will be provided to T&E or be made broadly available for additional research/analyses

- Offerors will be required to propose and implement the necessary safeguards to protect and maintain subjects' privacy, data, and other legal, ethical, and safety protections

- Safeguards and protections must be in place throughout all phases of the study, to include during data collection, movement, analysis, and storage

Sensors - Overview

MOSAIC will aim to bring together multimodal sensors to test whether unobtrusive, continuous, and passive sensing provide new or complementary information to traditional assessment techniques.

- For this program sensors are considered anything that can be used to measure a person or the environment
- Think about sensors that would have higher probability of good compliance and reliability
- Likely that sensors will have to be provided to subjects
- Able to be procured by an independent group at scale to use for 150-200 individuals OR provided by offeror to Government at this same scale to enable T&E testing

Sensors

Required	**Optional***		**Out of Scope****

• Mobile phone (Android or iOS)	• Mobile physiological sensors
	   • Bands
	   • Hats
	   • Glasses
	   • Clothes
		• Patches/tattoos
		• Tablet/Laptop
		• Social or other media
		• Other creative and/or novel sensors
			• Obtrusive or uncomfortable for persistent carrying / wear
			• Impractical for consistent use
			• Invasive
			• Mobile EEG

This list is just notional, offerors will not be limited to this set

* It is expected that the proposed use of sensors will need to be accompanied by a technical justification, to include theoretical/empirical support, how it meets other program requirements, and how it is expected to gather data that will be useful for measuring target predictors and dimensions.

** If the current form/factor of a sensor would consider it to be "out of scope" then it is expected that an offeror can propose that sensor, if they are able to demonstrate how the sensor would meet program requirements by Phase 2 of the program.

Sensors - Considerations

- Relatively easy to use;

- If worn or carried, should be comfortable enough to be imperceptible to the wearer/carrier throughout waking hours and during sleep;

- Persistent battery life, at least 24 hours of continuous use without recharging;

- Ergonomic design and size that does not obstruct normal movement, circulation, respiration, ingestion, digestion, neural function, social dynamics, or other activities and processes; and,

- Durability and robustness to some environmental conditions (e.g. temperature, moisture) and large impacts or forces.

Context - Overview

**MOSAIC seeks innovative solutions to sense and assess context to improve the interpretation of a subject's signals.**

Lisovich et al., 2010

CrowdSignals Aims to Create a Marketplace for Smartphone Sensor Data

MIT Media Lab
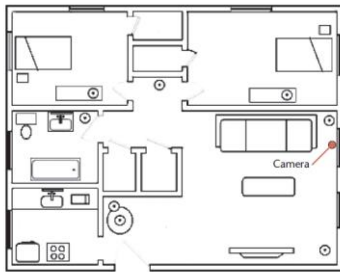
Low-Power Ambient Sensing in Smartphones for Continuous Semantic Localization

Sinziana Mazilu, Ulf Blanke, Alberto Calatroni, and Gerhard Tröster
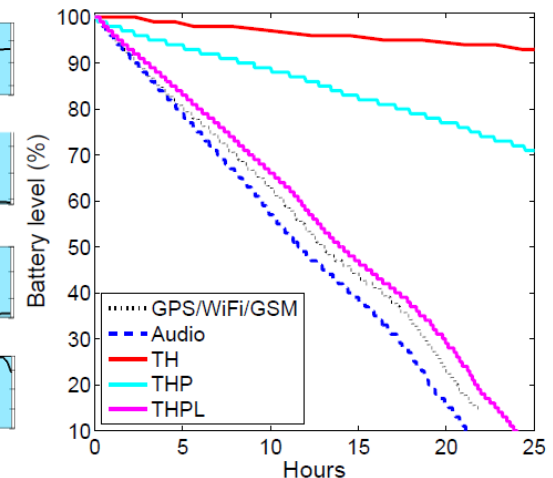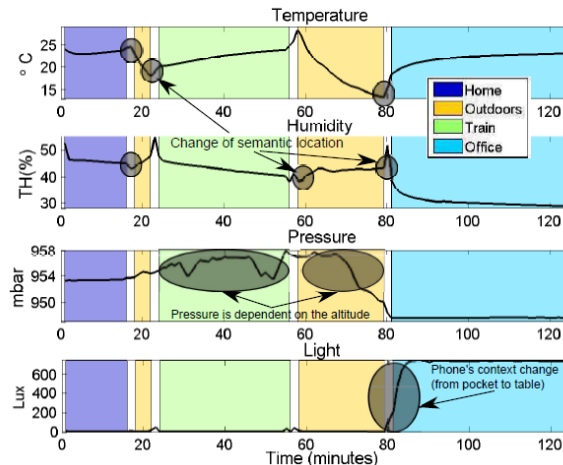
Context – Research Challenges

- Enable power hungry persistent & passive contextual sensing

- Take advantage of improving size, weight, power (SWaP) in sensors, chips, devices, and software

- Triage of ambient signals (informative or unhelpful)

- Distinguish similar, but different environments

- Solutions should not require <u>constant</u> GPS, wifi, etc.

Cleaning & Processing

MOSAIC seeks to develop capabilities to streamline and automate the cleaning and processing of multimodal, continuously collected data streams.

- Today this is one of the most time consuming steps and a big hindrance to using multimodal, continuously sampled data sources

- We want to move beyond manual correction and alignment or "boutique" one-off methods

- Offerors will be asked to propose methods and the development of tools to reduce the time to clean and process data with a pathway for how to automate this by the end of the program

- All methods and tools will be independently evaluated by the MOSAIC T&E team

  - Ideal solutions will not require a human-in-the-loop

  - Solutions that do require manual effort will have a high requirement for detailed documentation
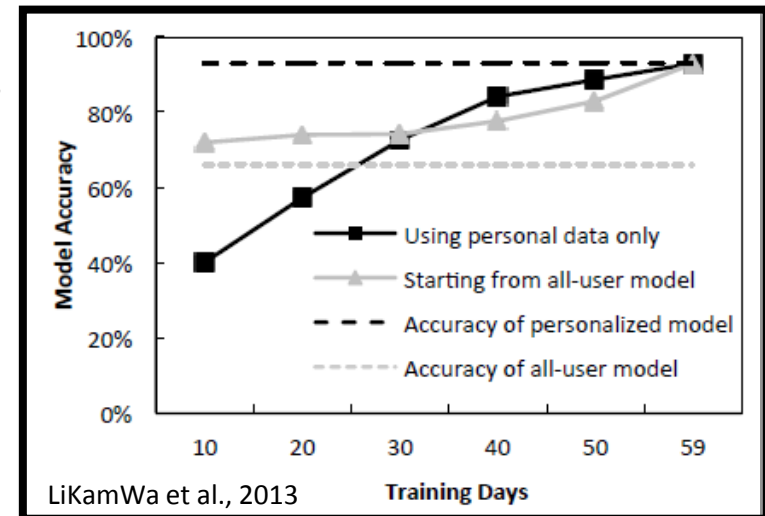
Modeling & Analysis

The program will advance modeling and analysis of personalized assessments over time and aided by the contextual information.

Research Challenges

- Develop a system that can be personalized to an individual

- Developing an accurate model without continual feedback from the user

- Developing a modeling approach that can evolve over time – we are not aiming to build a permanent model of an individual, but rather be flexible enough to detect shifts and then adapt once this has been detected

- Dealing with missing data

- Fusing data from multiple sources

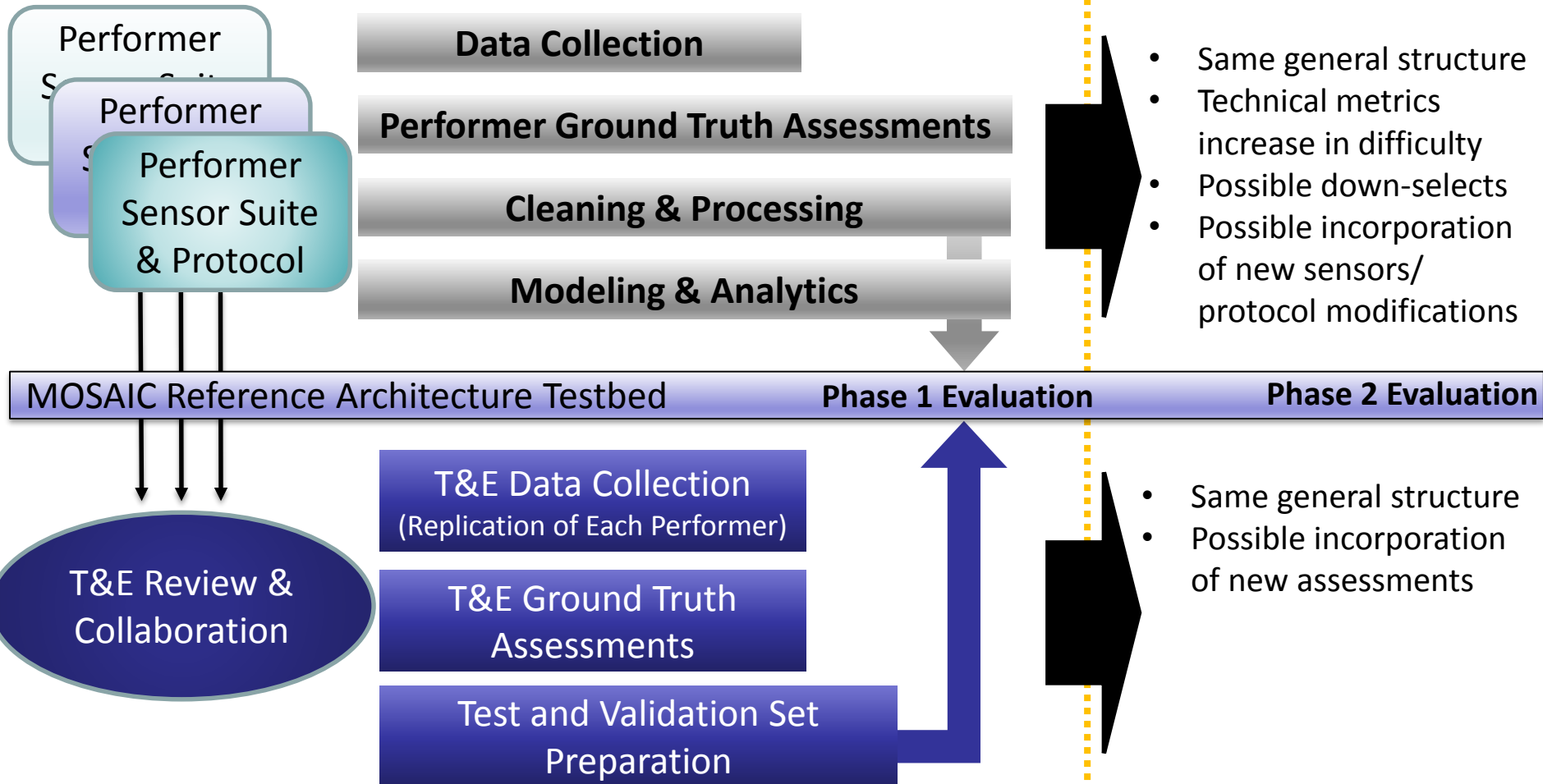- Ultimately real world solutions cannot require continuous labeling from subject

LiKamWa et al., 2013

Notional MOSAIC Flow Diagram

Phase 1

Phase 2

Performer Sensor Suite

Performer Sensor Suite

Performer Sensor Suite & Protocol

Data Collection

Performer Ground Truth Assessments

Cleaning & Processing

Modeling & Analytics

- Same general structure
- Technical metrics increase in difficulty
- Possible down-selects
- Possible incorporation of new sensors/ protocol modifications

MOSAIC Reference Architecture Testbed

Phase 1 Evaluation

Phase 2 Evaluation

T&E Data Collection
(Replication of Each Performer)

T&E Ground Truth Assessments

Test and Validation Set Preparation

T&E Review & Collaboration

- Same general structure
- Possible incorporation of new assessments
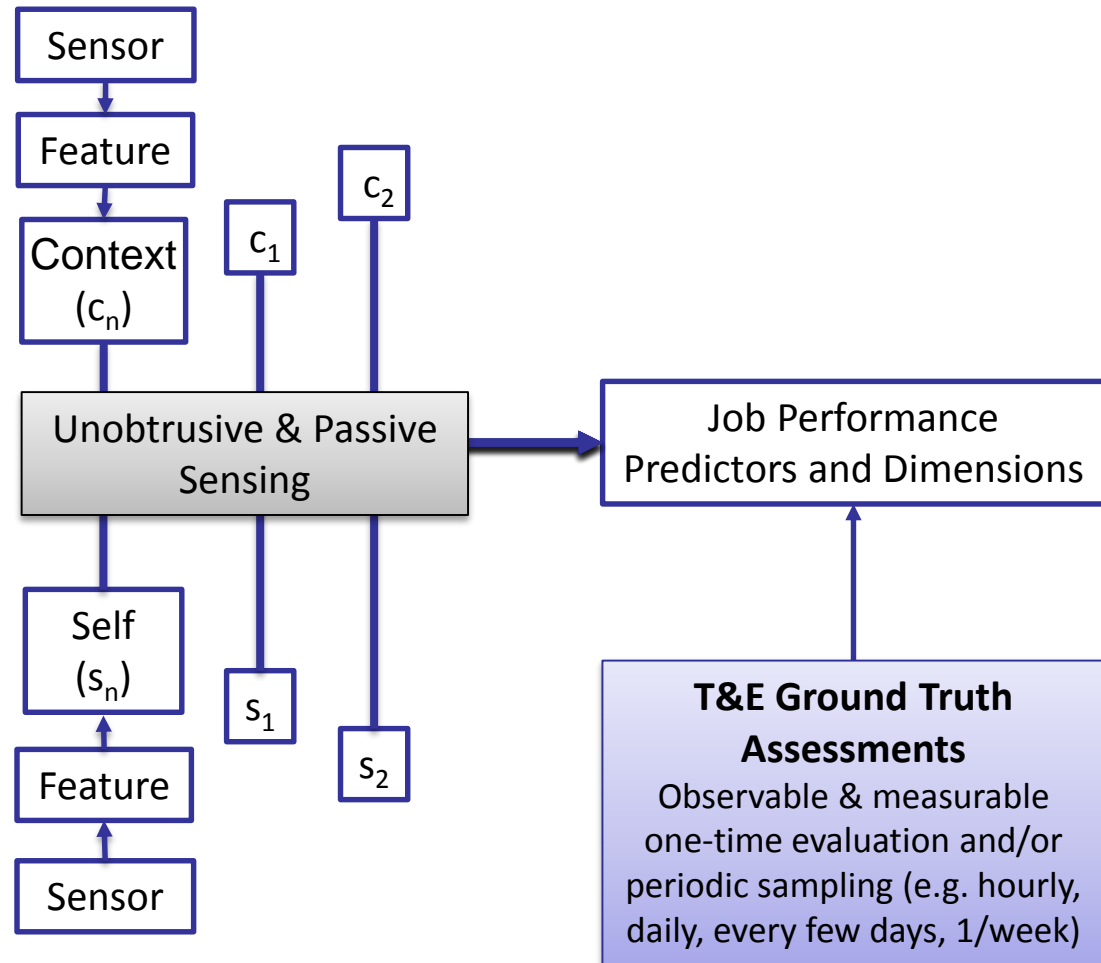
Test & Evaluation Overview

- MOSAIC Evaluations will occur once each phase.

- Performers will be evaluated on several metrics.

- Technical metrics will become progressively more challenging moving from Phase 1 to Phase 2.

- Each phase the evaluation will occur towards the end of the phase.

MOSAIC Test & Evaluation Schema

Context: Feature of the environment that can be objectively sensed and assessed (e.g. time, place, social or physical attributes).

Self: Feature of the target subject that can be objectively sensed and assessed.

Sensor

↓

Feature

↓

Context (c_n)

c_1

c_2

Unobtrusive & Passive Sensing → Job Performance Predictors and Dimensions

Self (s_n)

s_1

s_2

Feature

↑

Sensor

T&E Ground Truth Assessments
Observable & measurable one-time evaluation and/or periodic sampling (e.g. hourly, daily, every few days, 1/week)

Test & Evaluation Metrics

1 | **Can persistent, unobtrusive, and passive sensing be used to measure predictors and dimensions of job performance?**

2 | **Can persistent, unobtrusive, and passive sensing be used to detect changes in predictors and dimensions of job performance?**

3 | **Can the time to clean, process, model, and analyze the data be reduced?**

Metrics each phase are based on performers' results on T&E evaluations

The BAA will define specific targets for each metric, which may include:
- High classification accuracy and low false classification, as indicated by accepted measures of model classification accuracy (e.g., sensitivity, specificity, precision, recall)
- Feature selection and ranking (e.g. F score, mutual information)
- Incremental or predictive validity (e.g., effect size improvement, reduction in error)
- Model goodness-of-fit (e.g., likelihood ratio)

MOSAIC BAA Overview

Program Details

|	Phase 1	Phase 2
Structure	Competitive source selection through open BAA	Possible down selection from Phase 1 performers
Duration	21 months	21 months
Data	New data collection by performers on diverse set of subjects (in special cases possible use of existing datasets)	New data collection by performers and use of Phase 1 data, if relevant
T&E	Replication of Performer protocols; collection of independent datasets	Same as Phase 1
Evaluation	End of phase; will be used to establish performance on technical metrics	Same as Phase 1, metrics become progressively more challenging

Milestones & Waypoints

- **Milestones** are Government-defined progress metrics that must be met by the end of each phase

- **Waypoints** are offeror-defined, task-driven intermediate steps towards a milestone
    - Depending on an offeror's specific approach, progress towards a milestone is not expected to be linear in all areas
    - Waypoints are how the offeror clearly explains to the Government the quantitative and timely progress that must be made for their overall concept to meet the end-of-phase Milestones –performance against these waypoints is reviewed throughout program

- **Technical reviews and site visits** held at months throughout the program will quantify progress against the waypoints & assess whether course corrections are needed for success

Program Schedule
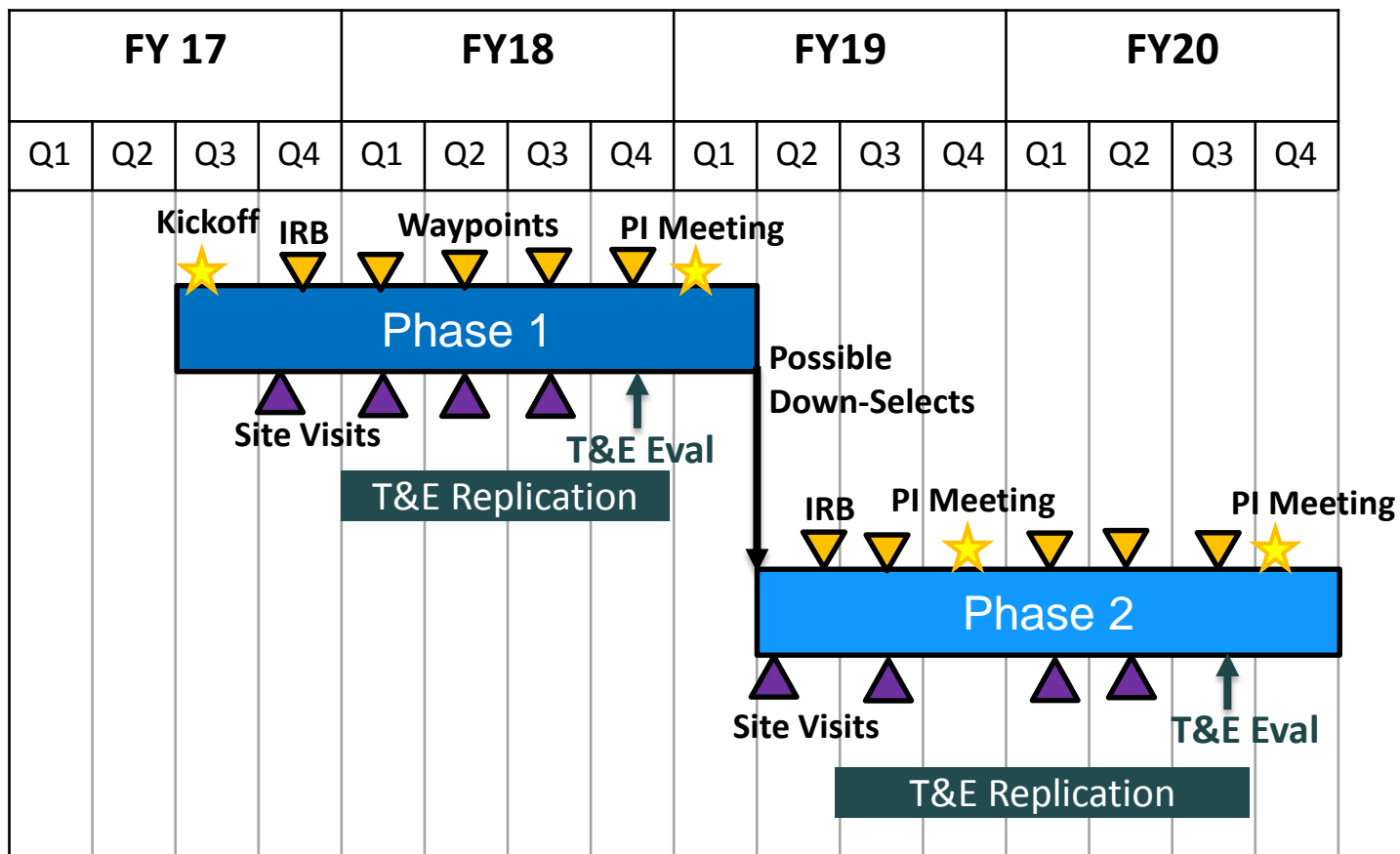
BAA Highlights

- MOSAIC is a research program not an operational activity, we are seeking innovative R&D to address several research challenges

- Proposals must address all technical areas in the BAA and innovation is encouraged in all areas

  1. Theoretical/empirical support for sensor approaches and data collection protocols

  2. Streamlined and automated methods for cleaning and processing multimodal data

  3. Personalized and temporally-rich modeling and analytics

- It is expected that the BAA will ask for a plan for assuring subject and data privacy protections throughout data collection, storage, and analysis.

BAA Highlights (cont.)

- Single BAA for Phases 1 and 2

- The Government anticipates that proposals submitted under this BAA will be unclassified

- Multiple awards are expected

- Foreign participants and/or individuals may participate to the extent that such participants comply with any necessary Non-Disclosure Agreements, Security Regulations, Export Control Laws and other governing statutes applicable under the circumstances

- Publications and presentations at conferences will be encouraged, but will require review by government prior to submission

MOSAIC Out-of-Scope

- Will **NOT** fund work on:
  - New device development
  - Improvements in SWaP components
  - Advances in radio technology
  - Advances in materials
  - Group level assessments
  - Medical/health apps
  - Medical, physical or mental health diagnostics
  - Biometric verification and identification
  - Use of blackbox sensors, processing pipelines, or algorithms

Additional Program Considerations

1. Devices & data collection

- Generally software, APIs and hardware platform will not be not limited, but may be negotiated between performers and Government to facilitate T&E

- Leverage open source solutions, where possible

- Device agnostic, with possible incorporation of new technology in Phase 2, and/or possible down-select of devices/sensors given Phase 1 results

- All de-identified datasets will be considered deliverables

Additional Program Considerations

2. Analysis

– Data sources, processing, analysis and results must be traceable and explainable, no blackbox solutions

– Will expect results to detail information about which sensors, set of sensors, and features were most predictive *and* those which were not

– May develop novel data visualization techniques

3. Reporting

– Details about the data collected including clear definitions of collection parameters, signal processing, feature selection, machine learning, etc.

– All source code and executables with a minimum of government purpose rights (GPR; unless listed and justified up front in proposal) and with the ability to independently run, verify, and modify, if necessary

Team Composition

Given the combination of technical challenges, we anticipate teams will possess expertise in:

- Social, behavioral, and cognitive sciences

- Industrial/Organizational psychology

- Multimodal sensing (software & hardware)

- Computer science, signal processing, computational modeling, algorithm development, machine learning

- Mathematics, statistics, psychometrics, data science

Summary

- MOSAIC seeks to develop and validate unobtrusive sensor-based methods to assess static and dynamic aspects of an individual's cognitive abilities and styles, interpersonal style, and well-being.

- The program aims to take advantage of multimodal sensors to develop personalized assessments.

- We are looking for well-executed, creative ideas for all technical areas.

- We expect teams will have a strong plan for working across team members to accomplish the program goals.

- The BAA supersedes anything presented or said at the Proposers' Day by IARPA.